

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
КАФЕДРА ТЕХНОЛОГИИ ПРОГРАММИРОВАНИЯ

Лескин Владимир Николаевич

Выпускная квалификационная работа бакалавра

Распознавание эмоций в видеопотоке

Направление 010300

Фундаментальная информатика и информационные технологии

Научный руководитель:

ст. преподаватель

Уланов А.В.

Санкт-Петербург  
2016

# Содержание

Введение.....	3
Постановка задачи .....	4
Обзор литературы .....	5
Глава 1. Обзор готовых решений и методов .....	6
1.1 Обзор существующих реализаций .....	6
1.1.1. FaceReader .....	6
1.1.2. EmoDetect .....	7
1.1.3. FaceSecurity .....	9
1.1.4. MS Oxford Project Emotion Recognition .....	11
1.2 Методы обнаружения лиц .....	12
1.3 Методы распознавания эмоций .....	15
1.3.1 Метод опорных точек .....	15
1.3.2 Метод Виолы-Джонса .....	17
Глава 2. Реализация .....	19
2.1 Основная концепция реализуемого проекта .....	19
2.2 Выбор облачной технологии .....	21
2.3 Выбор технологии создания .....	22
2.4 Результаты .....	26
Выводы.....	28
Заключение .....	30
Список литературы .....	31

## Введение

В современном мире, в век развития информационных технологий, где искусственный интеллект уже не кажется научной фантастикой, а является предметом работы многих учёных, появилось понятие компьютерного зрения. Если кратко, то компьютерное зрение — это теория и технология создания машин, которые могут производить обнаружение, отслеживание и классификацию объектов. Для того, чтобы взаимодействие машины и человека проходило наиболее комфортно и удобно для последнего, необходимо создать технологию, способную воспринимать сигналы, посылаемые ей человеком, будь то голосовые команды (как это продемонстрировали исследователи из университета Тафтса на конференции AI-HRI в ноябре 2015 года[1]), определённые невербальные жесты (технология Kinect) или сигналы мимики лица (проявление эмоций и чувств). Считывание человеческих эмоций может быть как единственным способом взаимодействия так и вспомогательным, например, для того чтобы следить за состоянием человека во время выполнения основной работы. Такой подход сделает анализ команд более точным, а взаимодействие будет проще.

Само распознавание человеческих эмоций может стать основной частью проекта по определению лжи, где в дополнение к биометрическим параметрам тела добавится информация о мимике, играющей важную роль при решении такой задачи. И даже если отойти от науки, можно найти применение этой задачи в развлекательных ресурсах, так система может запоминать моменты из фильмов, которые, судя по эмоциям, понравились зрителю больше всего, и подбирать подобный материал с учётом сформированных интересов.

## **Постановка задачи**

Целью данной работы является разработка приложения, определяющего эмоциональное состояние человека на основе его мимики, представленной во входном видеопотоке. Главной задачей приложения является распознавание черт лица, свойственных определённым человеческим эмоциям и вычисление конечного, понятного пользователю, результата.

Для достижения этой цели поставлены следующие шаги:

- 1) Рассмотреть существующие решения поставленной задачи и сравнить их достоинства и недостатки.
- 2) Провести сравнение алгоритмов решающих поставленную задачу и проанализировать их достоинства и недостатки. Выбрать наиболее подходящий, то есть тот, при котором будут минимизированы затраты ресурсов машины, на которой будет выполняться приложение, времени выполнения, финансовые затраты при разработке, и максимизирована скорость выполнения.
- 3) Рассмотреть существующие языки программирования, выбрать тот, который наиболее подходит для реализации приложения, то есть тот, при выборе которого достигаются оптимальные значения из пункта 2.
- 4) Выбор платформы и технологии создания, на основе анализа существующих инструментов и средств разработки.
- 5) Реализовать алгоритм и тестирование приложения. Нахождение и исправление ошибок.

## Обзор литературы

В книге “Компьютерное зрение. Современный подход” [17], написанной Дэвидом Форсайтом и Джином Понсом определяется понятие компьютерного зрения и даётся информация об этой области в целом, а также рассказывается о решаемых задачах, таких как распознавание, обнаружение объектов, восстановление изображений и т.д.

В [18][19] и [8] рассмотрены методы обнаружения объектов и распознавания эмоций на изображении, представлены их математические обоснования и применение в реальных задачах, схожих с поставленной. В работах [7][8] авторы рассматривают задачу распознавания лиц и представляют её возможное решение, используя методы описанные в [9][10][11].

При анализе существующих систем были просмотрены реализации [3][4][5][6], выявлены их особенности, плюсы и минусы, найдено их применение в реальной жизни и сформированы требования для реализации собственного приложения.

В качестве инструментов для разработки были проанализированы [12][13] и [15], рассмотрены реализованные алгоритмы и методы, помогающие решить поставленную задачу, а также проведено сравнение предоставляемых конфигураций виртуальных машин и стоимости предоставления доступа к облачным сервисам.

Автор [21] описывает технологии создания веб-приложений средствами облачной платформы Microsoft (Windows) Azure и платформы разработки .NET Framework, руководство по которой дано в [20] и [22].

# Глава 1. Обзор готовых решений и методов

В данной главе рассматриваются готовые решения поставленной задачи, которые были успешно разработаны и внедрены в широкое пользование, а также существующие алгоритмы и методы, с помощью которых достигается решение основной задачи.

## 1.1 Обзор существующих реализаций

### 1.1.1 **FaceReader** (Noldus Information Technology (Нидерланды))[3]

Данное приложение может верно распознавать по выражению лица такие эмоции, как «счастье», «грусть», «злоба», «удивление», «страх», «отвращение» и «спокойствие» (Рисунок 1). Кроме того, FaceReader способен определить возраст, пол и расу людей. Приложение не нуждается в обучении и дополнительной настройке.

В программе реализованы технологии компьютерного зрения. В частности, это метод Active Template, заключающийся в наложении деформируемого шаблона на изображение лица:

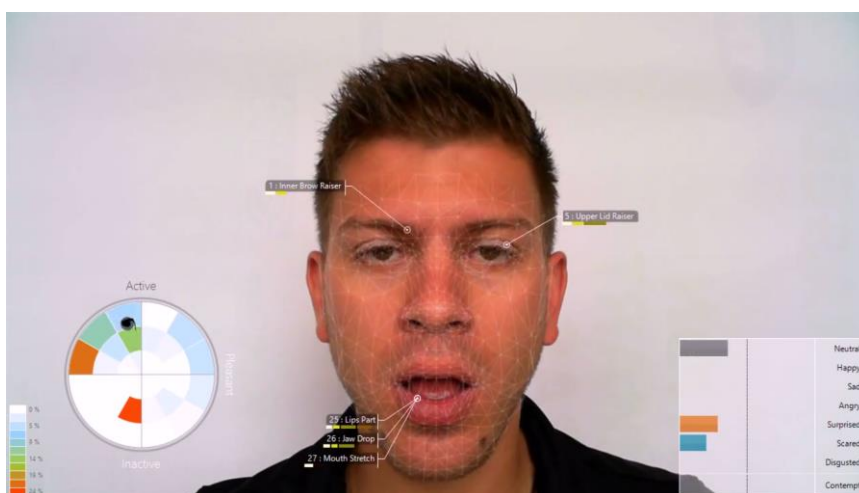


Рисунок 1- Интерфейс приложения FaceReader

Преимущества данной реализации:

- Средний процент распознавания эмоций равен 89% и не зависит от наклона и поворота лица плоскости.
- Программа работает с большим количеством форматов видеофайлов с кодеками MPEG1, MPEG2, DivX4, DivX5, DivX6, DV-AVI. При этом распознавание эмоций может происходить покадрово, либо полностью при просмотре всего видео. Приложение не ограничивается работой с загружаемым видео, FaceReader работает и со статичными изображениями и в реальном времени, считывая данные с камеры устройства пользователя.
- Возможности визуализации: наличие гистограмм, диаграмм, отображение накладываемой сетки, процентов выражаемых эмоций.

Недостатки:

- Процент распознавания эмоций и возраста сильно падает при работе с лицами детей возрастом меньше 5 лет.
- Распознавание может происходить неточно, если человек носит очки.
- Программа не определяет лица, повёрнутые в профиль.

### **1.1.2 EmoDetect (Нейроботикс, Россия)[4]**

Программное обеспечение EmoDetect позволяет определить психоэмоциональное состояние человека по выборке изображений (видео или набор файлов). Разработанный классификатор эмоций позволяет определить 6 базовых эмоций: радость, удивление, грусть, злость, страх, отвращение.

Определение эмоций происходит с помощью нейронных сетей: выполняется нахождение опорных точек на лице человека, и классификатор строит общую картину на основе их положения относительно друг друга.

Компания Нейроботикс разрабатывает и поставляет системы для исследования физиологии человека и животных и, используя эти знания, разрабатывает новые технологии в области биомеханики, биоуправления и сенсорики в робототехнике.



Рисунок 2 - Интерфейс приложения EmoDetect

Достоинства:

- Классификация эмоций тремя независимыми классификаторами: нейронная сеть, система решающих правил, классификатор по взвешенной сумме признаков.
- Построение графика динамического изменения интенсивности эмоций испытуемого во времени и формирование отчета о результатах обработки видео.
- Возможность записи видео-файлов с веб-камеры, анализ всех поддерживаемых операционной системой форматов



Недостатки:

- Программа не определяет лица, повернутые в профиль.

### 1.1.3 **FaceSecurity** ( Cognitec, Германия)[5]

Данная реализация сканирует поток входящих данных и, при обнаружении в кадре лиц, сравнивает их с заведённой базой данных на предмет наличия совпадений. В случае нахождения совпадений система отправляет уведомления операторам в реальном времени.

Данное приложение широко используется в системе безопасности европейских стран для выявления подозрительных лиц в общественных местах. Крупные компании же используют его для идентификации важных клиентов.

Рядовые фирмы используют данную разработку для распознавания лиц своих сотрудников (Рисунок 3) и выявления незарегистрированных посетителей, используя базы данных с фотографиями своих подчинённых. Также данная разработка помогает следователям идентифицировать лица в местах преступления по фотографии и видеонаблюдению путем сопоставления изображений лиц в хранилище агенств.

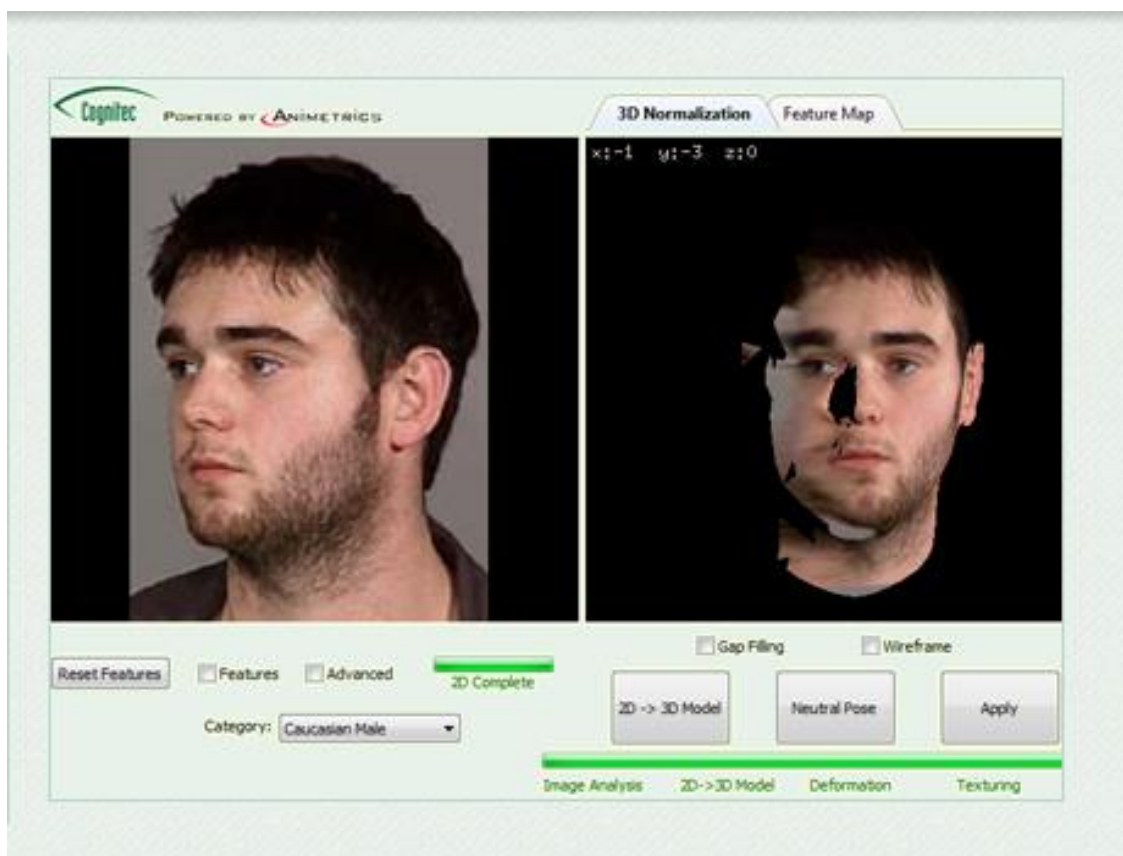


Рисунок 3 - Интерфейс приложения FaceSecurity

#### Достоинства:

- Работа в реальном времени, возможность слежки за обнаруженным лицом на нескольких потоках данных одновременно.
- В режиме реального времени лицо сравнивается с эталонами, занесёнными в базу данных.
- Регистрации при неподвижном изображении или при живом потоке видео в ручном и автоматическом режиме.
- Применение C++ API и Web Services API.
- Возможность масштабируемости в пределах «контрольного списка»: размер, количество видеопотоков и число видимых лиц на изображении.

#### Недостатки:

- Невозможность произведения вычислений для лиц, имеющих отклонение от фронтального положения более чем 15 градусов.

- Светочувствительность. Невозможность распознавания лица, при попадании на него теней и бликов.

#### 1.1.4 Microsoft Oxford Project Emotion Recognition (Microsoft, США)[6]

Принцип работы Emotion Recognition аналогичен предшественникам: машинные алгоритмы анализируют наличие лиц на загружаемой фотографии, после чего определяют по мимике предположительные эмоции. Учитывая тестовый режим работы этого инструмента, результаты часто получаются неточными и неожиданными. Предоставленное API в доступной форме даёт разработчикам всю мощь алгоритма машинного зрения и распознавания эмоций для использования в своих приложениях.



Рисунок 4 - Интерфейс демонстрационной версии приложения Emotion Recognition

Достоинства:

- Вычисление процентов выражаемых эмоций
- Доступно Emotion API, которое может быть использовано в различных приложениях.

Недостатки:

- Определение эмоций только на статичных изображениях.
- Программа не распознаёт лица, расположенные под большим углом ( $>40^\circ$  градусов)
- На данный момент доступна только демонстрационная версия.

## 1.2 Методы обнаружения лиц

**Метод Виолы-Джонса** (*Viola-Jones object detection*) - алгоритм обнаружения объектов на изображениях в реальном времени, разработанный в 2001 году Полом Виолой и Майклом Джонсом. Несмотря на то, что главной задачей при разработке этого метода было нахождение лиц, этот метод используется и для распознавания других различных предметов. Существует множество реализаций, в том числе в составе библиотеки компьютерного зрения OpenCV (функция `cvHaarDetectObjects()`).

Признаки, применяемые алгоритмом, используют суммирование пикселей из прямоугольных регионов. Сами признаки несколько напоминают признаки Хаара, которые также используются для поиска объектов на изображениях. Однако признаки, которые использовали Виола и Джонс сложнее и содержат в себе более одной прямоугольной области (Рисунок 5).

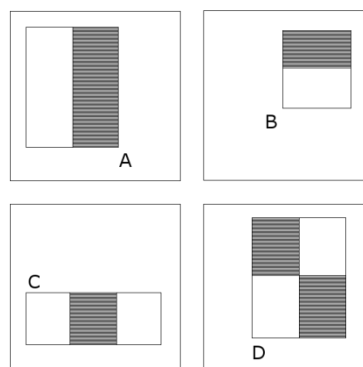


Рисунок 5 – Признаки, использованные Виолой и Джонсом

На рисунке 5 показано 4 различных типа признаков. Величина каждого признака вычисляется как сумма пикселей в белых прямоугольниках, из которой вычитается сумма пикселей в черных областях. Прямоугольные признаки более примитивны чем поворачивающийся фильтр (steerable filter), и, несмотря на то, что они чувствительны к вертикальным и горизонтальным особенностям изображений, результат их поиска более груб. Однако, их преимуществом по сравнению с более точными вариантами заключается в том, что при хранении изображения в интегральном формате<sup>1</sup>, проверка прямоугольного признака на конкретной позиции проводится за константное время. Каждая прямоугольная область в используемых признаках всегда смежна с другим прямоугольником, поэтому расчет признака с 2 прямоугольниками состоит из 6 обращений в интегральный массив, для признака с 3 прямоугольниками - из 8, с 4 прямоугольниками - из 9. Высокая скорость обсчета признака не компенсирует значительное количество различных возможных признаков. При стандартном размере признака в 24x24 пикселя, возможно 162 тысячи разных признака и производить расчёт каждого слишком долго. Поэтому в алгоритме Виолы-Джонса используется вариация алгоритма обучения AdaBoost[2], как для выбора признаков, так и для настройки классификаторов.

Преимущества данного алгоритма:

- Быстрое вычисление значений признаков
- Эффективный набор признаков
- Вместо масштабирования изображения, происходит масштабирование признаков

Таким образом, алгоритм находит лица с высокой точностью, то есть низким процентом ложных срабатываний и высоким процентом верного обнаружения.

---

<sup>1</sup> integral image, в каждом пикселе изображения записана сумма всех пикселей левее и выше данного

К недостаткам же относится невозможность распознать искомый объект при его повороте на 45 и более градусов, чувствительность к углу падения света, а также длительное время обучения классификаторов.

**Eigenfaces** (собственные лица) - название совокупности собственных векторов, используемых в задаче компьютерного зрения распознавания человеческих лиц. Подход с использованием такого набора для распознавания был разработан L. Sirovich и M. Kirby в марте 1987 [7] и применён Мэтью Терком и Алексом Пентландом (Matthew Turk and Alex Pentland) в классификации лиц [8].

Для того, чтобы найти, так называемые собственные лица, используется метод главных компонент [9]. Как и для любого другого алгоритма с машинным обучением, Eigenface требует обучающую выборку, представляющую собой изображения лиц, которые необходимо распознать. После этого, порядок действия следующий:

1. Имеющаяся обучающая выборка из  $N$  изображений размера  $n \times m$  пикселей представляется в виде матрицы размера  $nm \times N$ , называемой пространством лиц. Для этого "вытягивается" каждое изображение в вектор длиной  $nm$ , то есть все строки пикселей записываются друг за другом (конкатенация). Каждый вектор образует столбец в матрице.
2. Вычисление собственных векторов этой матрицы и упорядочивание их в порядке убывания соответствующих им собственных значений. Это "собственные лица".
3. Правило сломанной кости[10] позволяет найти оптимальное число  $M$  собственных векторов, хорошо описывающих выборку, но в тоже время не избыточное.

Набор из  $M$  значений образует вектор в пространстве лиц. Для распознавания лица на изображении нужно найти соответствующий этому

изображению вектор в пространстве лиц и определить к какому вектору из обучающей выборки он ближе всего. Для оценки расстояния целесообразно использовать дистанцию Махаланобиса [11].

Данный алгоритм больше подходит для идентификации человека по фотографии, то есть сравнение его лица на фотографии с изображением из эталонной базы.

## 1.3 Методы распознавания эмоций

### 1.3.1 Поиск ключевых точек.

Для того, чтобы распознать эмоции необходимо иметь информацию о, так называемом, состоянии рассматриваемого объекта. Таким образом первый шаг - это определить точное положение опорных точек лица (Рисунок 6). Для начала необходимо дать определение нескольким понятиям: двигательная единица - величина, принимающая значение от 0 до 1, определяющая то, насколько далеко находится соответствующая опорная точка от линии глаз или нормали к этой прямой, если значение двигательной единицы  $> 0$ , то такая единица называется активированной (Рисунок 7), нейтральным расстоянием называется величина, обозначающая среднее расстояние от опорной точки до нужной прямой при нейтральном выражении лица, то есть при таком, что коэффициенты проявления каждой эмоции, принимающие значения от 0 до 1, не превышали 0.26, максимальным - расстояние от прямой до точки, относящейся к двигательной единице при проявлении связанной с ней той или иной эмоции.

В данном методе таких точек 10: по 2 точки для бровей, зрачки, углы рта, верхняя и нижняя губы (рисунок 6).



Рисунок 6 - Расположение опорных точек

Для каждой из точек устанавливается область интереса (Region of Interest) на рассматриваемом изображении, чтобы уменьшить район поиска, снизив время нахождения, обычно эти области находятся с помощью классификаторов обнаружения лиц, глаз, рта и т.д. Нахождение самих точек в выделенных областях происходит двумя способами: для бровей применяется адаптивный контрастный фильтр, для которого задаётся минимальное и максимальное значения серого цвета для обнаружения границ брови, которая впоследствии представляется набором соединённых точек или полигоном. Для поиска зрачков обычно используется обученный классификатор, определяющий точное положение глаз и зрачков на изображении. Чтобы определить положение оставшихся четырёх точек используются оба подхода: первым делом классификаторы разбивают область поиска на 4 части, а затем применяют адаптивный контрастный фильтр для нахождения граничных точек.

Для того, чтобы на основе найденных точек определить изображаемые эмоции используется система кодирования лицевых движений (Facial Action Coding System), разработанная Полом Экманом и Уоллесом Фризеном [15]. После определения всех ключевых точек вычисляются значения, так называемых, двигательных единиц (action units). Для этого необходимо задать нейтральное и максимальное расстояния от ключевых точек до прямой, проходящей через координаты расположения зрачков или нормали к этой



прямой. Это можно сделать путём анализа нескольких изображений, найдя средние значения этих расстояний. После этого на заданном изображении находится расстояние от ключевых точек до нужной прямой и сравнивается с нейтральным и максимальным. Поскольку размеры изображений могут оказаться разными, то за 1 берётся расстояние между зрачками. Если найденное расстояние больше или равно максимальному, то значение двигательной единицы равно 1, если оно меньше нейтрального, то значение равно 0, а если находится между нейтральным и максимальным расстоянием, то значение вычисляется линейной интерполяцией.

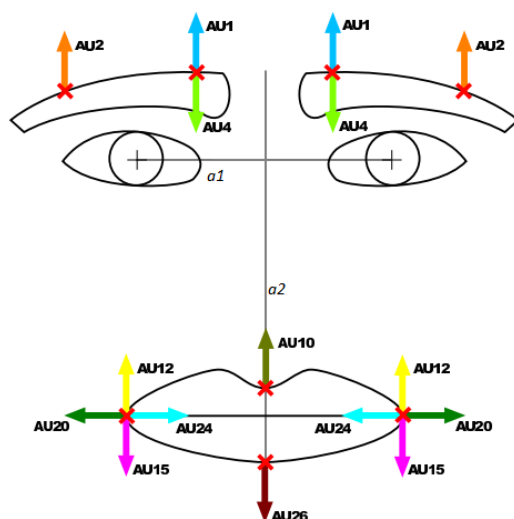


Рисунок 7 - Двигательные единицы. Стрелками обозначены направления, в котором ключевые точки должны двигаться, чтобы активировать двигательные единицы

После нахождения всех значений, для вычисления конкретных эмоций необходимо сложить полученные значения, которые влияют на проявление этой эмоции и разделить на количество взятых единиц. Таким образом получится число от 0 до 1, показывающее степень выражения вычисляемой эмоции.

### 1.3.2 Метод Виолы-Джонса.

Рассмотренный ранее алгоритм Виолы-Джонса также можно применить и к распознаванию эмоций человека. Для этого необходимо, чтобы система умела находить на изображении несколько вариантов частей лица, например, рот

человека, при проявлении счастья, грусти, удивления, то же самое с глазами, бровями, носом. Очевидно, что чем больше эмоций должна распознавать система, тем больше обученных каскадов потребуется. Это означает, что анализировать один кадр система будет долго, в зависимости от аппаратного обеспечения это может занимать несколько секунд, что в условиях поставленной задачи неприемлемо. Данный метод хорошо работает для распознавания эмоций на изображениях, но не в видеопотоке. К тому же обучение каскадов занимает много времени, от нескольких дней до недель, и требует большой обучающей выборки (не менее 1000 положительных и 1000 негативных изображений) для получения низких процентов ложных срабатываний и высоких процентов обнаружения.

## Глава 2. Реализация.

В данной главе рассматриваются и сравниваются технологии создания приложения, а также проводится выбор на основе их плюсов и поставленных условий реализации.

Для реализации был выбран алгоритм Виолы-Джонса для обнаружения лиц в кадре, поскольку данный метод имеет низкое количество ложных срабатываний и высокую точность обнаружения искомых объектов. Вычисление признаков происходит быстрее чем у методов со схожим принципом работы [23]. Метод *eigenfaces* в основном используется для нахождения лиц, схожих с заданным, и не используется только для их обнаружения.

Однако данный алгоритм требует больших временных ресурсов на обучение каскадов, число которых увеличивается при увеличении количества распознаваемых эмоций, также необходимость обнаружения большого числа искомых объектов ведёт к медленной обработке изображения. Поэтому для распознавания эмоций был выбран алгоритм, основанный на нахождении опорных точек лица.

### 2.1 Основная концепция реализуемого проекта.

Главными требованиями к конечному приложению являются:

- Возможность предоставления данных для их обработки и вывода понятных пользователю результатов.
- Обработка минимум 5 лиц в одном кадре.
- Вывод максимально адекватных результатов (процент правильного распознавания доминирующей эмоции при финальном тестировании не должен быть ниже 70%).
- Устойчивость к потерям интересующих объектов из кадра в силу подвижности объектов в видео.

На основе представленных требований было принято решение разработать приложение, работающее по следующему алгоритму:

1. Приложение получает доступ к веб-камере пользователя.
2. Инициализация необходимых объектов для последующей работы.
3. Считывание кадра из видеопотока с камеры и отправка запроса на сервер.
4. Обнаружение лиц, вычисление координат интересующей области. В случае, если лицо в кадре не обнаружено, вывод соответствующей информации клиенту и переход к пункту 6.
5. Работа с найденной областью, предоставление результатов на клиентскую сторону.
6. Если камера не отключена, возврат к пункту 3.

При разработке приложения, работающего с видеопотоками и изображениями было принято решение обрабатывать полученную информацию на сервере, поскольку разница в программном обеспечении пользователей может негативно сказаться на работе приложения, поскольку оно не будет должным образом оптимизировано под ту или иную конфигурацию системы. Чтобы избежать таких возможных проблем, имеет смысл использовать облачные технологии, которые предоставляют возможность проводить вычисления, в том числе и параллельные, на предоставленных серверах, тем самым не ориентируясь на конкретные конфигурации систем конечных пользователей, а проводить все вычисления на сервере, передавая пользователям только результат.

В качестве реализации было решено разработать веб-приложение, которое будет иметь доступ к веб-камере пользователя, тем самым получая поток входных данных, затем обрабатывать полученную информацию на сервере и возвращать результат на страницу в браузере пользователя. Такое решение позволит эффективно использовать предоставленные ресурсы, выделяя нужное их количество при необходимости и освобождая при ненадобности. Доступ к

таким приложениям возможно осуществлять с любого устройства, независимо от его аппаратного обеспечения. Такая реализация может быть использована для разработки расширений браузеров или плагинов для приложений с функцией видеосвязи, например Skype, WhatsApp, Viber.

## 2.2 Выбор облачной технологии

На сегодняшний день можно выделить три наиболее популярных компании, предоставляющих доступ к облачным технологиям: Microsoft, Amazon и Google.

1. Google App Engine (последняя версия 1.9.35, 24 марта 2015 года) - сервис размещения сайтов и веб-приложений, представленный в апреле 2008 года. Сервис предлагает множество инструментов для работы с аккаунтами Google и хранилищем данных BigTable. Особенности:

- Использование PHP, Go, Python или Java для разработки приложений.
- Использование проприетарной базы данных BigTable, что дополнительно потребует оптимизации приложений.
- Поддержка автомасштабирования.
- Бесплатная<sup>2</sup> пробная версия на 60 дней.

2. Amazon Web Services - набор сервисов облачных вычислений, представленный в 2006 году. Центральной частью этого набора является сервис Elastic Compute Cloud (EC2), который предоставляет возможность хранить и выполнять приложения на серверах Amazon. Особенности:

- Оплата только за активные сервера.
- Позволяет выбрать географическое местоположение сервера, что позволяет снизить латентность.
- Минимальная стоимость - 325\$ [16].

---

<sup>2</sup> Предоставление 300 долларов США на использование сервисов Google Cloud Platform сроком на 60 дней

3. Microsoft Azure - облачная платформа и инфраструктура, представленная в феврале 2010 года. Предоставляет возможность разработки и выполнения приложения на выделенных серверах. Работоспособность платформы обеспечивается с помощью сети дата-центров, распределённых по всему миру. Особенности:

- Оплата только потреблённых ресурсов.
- Возможность как автомасштабирования, так и программного.
- Поддержка большого количества технологий и языков программирования, например ASP.NET, Java, PHP и т.д.
- Поддержка пользовательских сертификатов SSL.
- Бесплатная конфигурация ресурсов.
- Была признана самой быстрой облачной платформой [14].

## 2.3 Выбор технологии создания.

Для реализации приложения выбрана библиотека компьютерного зрения с открытым исходным кодом OpenCV [12], а также обёртка для этой библиотеки EmguCV [13], позволяющая использовать классы и методы исходной библиотеки в .NET совместимых языках программирования. OpenCV содержит реализованные функции поиска лица на изображении с использованием признаков Хаара.

Microsoft Azure позволяет разрабатывать серверную часть веб-сайтов и приложений с использованием технологии ASP.NET, разработанной компанией Microsoft и использующей все возможности платформы .NET Framework. Это означает, что код приложения можно писать на любом языке программирования, поддерживаемом платформой .NET, например C#, Visual Basic .NET, JScript.NET. В сравнении со скриптовыми технологиями, данное решение также имеет преимущество в скорости за счёт того, что при первом

обращении код компилируется и помещается в специальный кэш, а при последующих только исполняется.

В качестве языка программирования для разработки серверной части приложения выбран объектно-ориентированный язык C#, поскольку во первых: данный язык был создан как язык разработки приложений для платформы .NET Framework. А во вторых выбранная библиотека OpenCV имеет обёртку для этого языка программирования. Пользовательский интерфейс представляет собой веб-страницу.

Взаимодействие между клиентом и сервером обеспечивается с помощью асинхронных ајах запросов, которые позволяют обмениваться данными не перезагружая страницу в браузере.

Поскольку в данной работе был выбран метод Виолы-Джонса для обнаружения лица и его частей и метод ключевых точек для распознавания эмоций, а также облачная платформа Microsoft Azure для хранения и выполнения приложения, то для успешной реализации проекта необходимо выполнить следующее:

1. С помощью классификатора, поставляемого вместе с библиотекой OpenCV, обучить каскады для обнаружения на кадре лица, рта, глаз.
2. Реализовать серверную часть приложения, где будет происходить анализ кадра и выявление ключевых точек методами, описанными ранее, а также формирование понятного пользователю вывода, описывающего процент выражаемой эмоции.
3. Реализовать клиентскую часть приложения: обеспечить получение доступа к камере пользователя и возможность считывать поток видеоданных.
4. Обеспечить взаимодействие между клиентской и серверной частями приложения.
5. Развернуть приложение с помощью облачной платформы Microsoft Azure.

Обучение каскадов происходит с помощью программы из библиотеки OpenCV `opencv_traincascade`. Ей на вход подаётся обучающая выборка, включающая набор “положительных” и “негативных” изображений. “Положительные” - это изображения, которые содержат предмет поиска. Текстовый файл с перечислением всех таких изображений с координатами искомого объекта передаётся программе `opencv_createsamples`, которая на выходе предоставляет файл с расширением `.vec`, в котором содержится информация о выборке “положительных” изображений. “Негативные” содержат только различные фоны, на которых не должен присутствовать искомый предмет. После завершения работы программы обучения каскада создаётся файл с расширением `.xml`, куда заносится вся полученная информация. Это файл может быть применён для поиска искомого объекта на изображении.

Серверная часть реализуемого проекта будет написана с помощью технологии создания веб-приложений ASP.NET на языке программирования C#. Основной метод класса называется `UploadImage` и принимает в качестве входного аргумента строку в формате `base64`<sup>3</sup>, в которой содержится информация о передаваемом изображении, к этому изображению применяется поиск лиц с применением полученного каскада, и далее рассматривается только область полученного прямоугольника с найденным лицом. Для поиска оставшихся частей лица сначала задаётся область интереса на полученном прямоугольнике, например для поиска глаз рассматривается только верхняя часть лица, а для поиска рта - нижняя. Для вычисления координат опорных точек, к выделенной области применяется алгоритм Кэнни для нахождения границ на изображении, который выполняется в 5 шагов:

1. Сглаживание. На этом этапе оператор Кэнни для удаления шумов использует фильтр Гаусса. Уравнение для ядра размера  $(2k + 1) \times (2k + 1)$  этого фильтра выглядит следующим образом:

---

<sup>3</sup> Base64 - формат представления бинарных данных в виде строки с кодировкой ASCII.



$$H_{i,j} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right); 1 \leq i, j \leq (2k+1),$$

где  $\sigma$  - среднеквадратическое отклонение.

2. Поиск градиентов. Оператор Кэнни отмечает границы там, где градиент изображения принимает максимальное значение, и поскольку направление вектора градиента может быть различным, то угол его направления округляется до одного из четырёх значений:  $0^\circ, 45^\circ, 90^\circ$  или  $135^\circ$
3. Подавление немаксимумов. Граничными пикселями являются только те, в которых достигается локальный максимум градиента в его направлении.
4. Двойная пороговая фильтрация. Для этого устанавливаются 2 пороговых значения. Чем меньше порог, тем больше границ найдёт алгоритм, однако повышается восприимчивость к шуму на изображении, однако если задать слишком высокий уровень, то найденная граница может иметь разрывы в слабых местах. Если значение цвета пикселя выше верхней границы, граница считается достоверной, если ниже – пиксель подавляется, точки со значением, попадающим в диапазон между порогов, принимают фиксированное среднее значение
5. Трассировка области неоднозначности. Выделение пикселей и определение их как граничных или подавленных. Граничными становятся пиксели, которые соединены хотя бы в одном из восьми направлений с граничными пикселями.

Выделив контуры нужной части лица, нахождение искомых координат сводится к поиску точек, располагающихся левее, правее, выше или ниже всех остальных. После нахождения всех опорных точек вычисляется расстояние от них до линии глаз или перпендикулярной ей прямой, проходящей через её центр. Полученные данные соотносятся с таблицей нейтральных и максимальных расстояний, и вычисляются конечные коэффициенты

выражаемых эмоций. Для того, чтобы составить таблицу максимальных и нейтральных расстояний для двигательных единиц было проанализировано 140 изображений (по 20 на каждую из 6 эмоций и 20 на нейтральное выражение лица).

Дв.ед.	Нейтр	Макс	Дв.ед.	Нейтр	Макс.	Дв.ед.	Нейтр	Макс.
1	0.270	0.441	10	0.982	0.875	20	0.354	0.498
2	0.283	0.441	12	1.056	0.902	24	0.354	0.154
4	0.270	0.142	15	1.056	1.111	26	1.223	1.714

Таблица 1 - Максимальные и нейтральные расстояния двигательных единиц

На выход метод подаёт текстовую информацию, понятную пользователю, которая содержит информацию о степени проявления эмоций.

## 2.4 Результаты

В результате, используя преимущества облачной платформы Microsoft Azure, возможности программной платформы .NET Framework и технологии создания веб-сервисов ASP.NET, было разработано приложение, серверная часть которого написана с использованием языка программирования C# (Рисунок 8). Исходный код приложения находится в свободном доступе и размещён в веб-сервисе GitHub [24], служащий для хранения и совместной разработки проектов и использующий систему контроля версий Git [25]. Само приложение использует поддомен azurewebsites.net и размещено с именем CameraCaptureLeskinV [26].

Реализация удовлетворяет всем поставленным требованиям и обеспечивает:

1. Бесперебойный обмен данными между клиентом и сервером. Устойчивость к быстрой смене кадров и отсутствию лица в видеопотоке.

2. Решение поставленной задачи распознавания эмоций человека на основе его мимики и выражения лица.
3. Предоставление понятной пользователю информации о результатах, полученных в ходе вычислений и анализа данных.
4. Осуществление доступа к приложению через браузер устройства с веб-камерой.

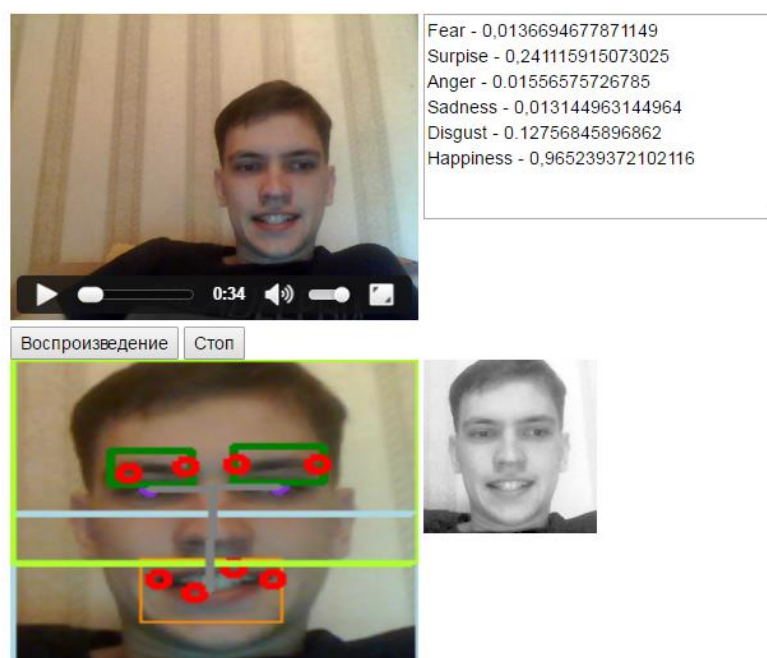


Рисунок 8 - Интерфейс приложения

В качестве вспомогательной информации выводится изображение найденного лица с отмеченными опорными точками. Если по какой-либо причине координаты некоторых точек не были вычислены, а остальные слабо изменили свое положение или не изменили вовсе, то пропущенным координатам присваиваются значения, вычисленные при анализе предыдущего кадра.

## Выводы

Метод Виолы-Джонса для нахождения лиц показал хорошие результаты: при тестировании использовалось 100 кадров со 150 лицами, 17 раз алгоритм совершил ошибку первого рода (не обнаружил лицо там, где оно было) и 15 раз - ошибку второго рода (обнаружил лицо там, где его нет). Группа изображений, при анализе которых были допущены ошибки, в большей степени состояла из изображений с наклонённой или повёрнутой головой, а также кадров со слишком тёмным или ярким освещением. Метод для распознавания эмоций, основанный на обнаружении опорных точек проявил себя хуже. Незначительные ошибки локализации одной точки имеют большое влияние на вычисляемое значение двигательной единицы, которое приводит к снижению точности распознавания. Тесты показывают, что эмоции счастье, удивление и печаль могут быть правильно обнаружены более чем в 70% случаев. Были проведены тесты с использованием 20 лиц на каждую эмоцию, хуже всего распознавалась эмоция отвращение - 60%, лучше всего приложение проявило себя при распознавании счастья - 90%. (Рисунок 9).



Рисунок 9 - Количество верных распознаваний эмоций

Для повышения качества распознавания возможны следующие варианты: увеличение числа опорных точек, что позволит снизить большое влияние малых

отклонений на конечный результат, или усложнение правил принятия решений, которые учитывают не только значения двигательных единиц, но и взаимное расположение опорных точек.

## Заключение

В данной работе была поставлена задача распознавания эмоций в видеопотоке и изучена её актуальность, рассмотрены существующие реализации и сравнение их преимуществ и недостатков. Проведён обзор алгоритмов поиска объектов, в частности лиц, на изображении. На основании этого обзора выбран методы Виолы-Джонса для обнаружения лиц в кадре и метод на основе ключевых точек для распознавания эмоций как приоритетные в реализации приложения решающего поставленную задачу. Во время работы с данным методом были обучены каскады, предназначенные для нахождения конкретных частей лица на изображении, а также успешно реализовано нахождение координат опорных точек, на основе которых строится конечный результат. Выявлены проблемы работы с данным алгоритмом и представлены их возможные решения

## Список Литературы

1. Artificial Intelligence and Human-Robot Interaction // AAAI Fall Symposium Series. URL: <http://ai-hri.github.io/> (дата обращения 12.12.2015).
2. AdaBoost // Wikipedia The Free Encyclopedia URL: <https://ru.wikipedia.org/wiki/AdaBoost> (дата обращения 12.12.2015).
3. Facial expression recognition software FaceReader // Noldus - Innovative solutions for behavioral research. URL: <http://www.noldus.com/human-behavior-research/products/facereader> (дата обращения 12.12.2015).
4. Нейроботикс - EmoDetect // Нейроботикс. URL: <http://neurobotics.ru/robotics/robotic-software/emodetect> (дата обращения 12.12.2015).
5. Products - Cognitec // The face recognition company - Cognitec. URL: <http://www.cognitec.com/products.html> (дата обращения 12.12.2015).
6. Microsoft Cognitive Services - Emotion API // Microsoft. URL: <https://www.projectoxford.ai/demo/Emotion> (дата обращения 12.12.2015).
7. L. Sirovich, M. Kirby. Low dimensional procedure for characterization of human faces//Journal of the Optical Society of America A, 1987, Vol. 4, P. 519.
8. M. Turk and A. Pentland. Face recognition using eigenfaces // Proc. IEEE Conference on Computer Vision and Pattern Recognition. 1991, P. 586–591.
9. Метод главных компонент // Wikipedia The Free Encyclopedia. URL: [https://ru.wikipedia.org/wiki/Метод\\_главных\\_компонент](https://ru.wikipedia.org/wiki/Метод_главных_компонент) (дата обращения 12.12.2015).
10. Правило сломанной кости. Метод главных компонент // Wikipedia The Free Encyclopedia. URL: [https://ru.wikipedia.org/wiki/Метод\\_главных\\_компонент#Оценка числа главных компонент по правилу сломанной трости](https://ru.wikipedia.org/wiki/Метод_главных_компонент#Оценка_числа_главных_компонент_по_правилу_сломанной_трости) (дата обращения 12.12.2015).

11. Расстояние Махаланобиса // Wikipedia The Free Encyclopedia. URL: [https://ru.wikipedia.org/wiki/Расстояние\\_Махаланобиса](https://ru.wikipedia.org/wiki/Расстояние_Махаланобиса) (дата обращения 12.12.2015).
12. About OpenCV // OpenCV. URL: <http://opencv.org/about.html> (дата обращения 12.12.2015).
13. Emgu CV: OpenCV in .NET // Emgu CV: OpenCV in .NET. URL: [http://www.emgu.com/wiki/index.php/Main\\_Page](http://www.emgu.com/wiki/index.php/Main_Page) (дата обращения 11.02.2016).
14. Windows Azure beats Amazon EC2, Google App Engine in cloud speed test // ARS Technica. URL: <http://arstechnica.com/business/news/2011/10/windows-azure-faster-than-amazon-ec2-and-google-app-engine-in-yearlong-cloud-speed-test.ars> (дата обращения 20.03.2016).
15. Система кодирования лицевых движений // Wikipedia The Free Encyclopedia. URL: [https://ru.wikipedia.org/wiki/Система\\_кодирования\\_лицевых\\_движений](https://ru.wikipedia.org/wiki/Система_кодирования_лицевых_движений) (дата обращения 15.02.2016).
16. Цены на инстансы EC2 // Amazon Web Services. URL: <https://aws.amazon.com/ru/ec2/pricing/> (дата обращения 20.03.2016).
17. Дэвид Форсайт, Жан Понс. Компьютерное зрение. Современный подход. 2004, 928 с.
18. Samuel Strupp, Norbert Schmitz, and Karsten Berns. Visual-Based Emotion Detection for Natural Man-Machine Interaction // KI 2008: Advances in Artificial Intelligence. 2008, Vol. 5243, P.356 -363.
19. Paul Viola, Michael J. Jones. "Robust Real-time Object Detection". 2001.
20. Руководство по разработке для .NET Framework // MSDN - сеть разработчиков Microsoft. URL: [https://msdn.microsoft.com/ru-ru/library/hh156542\(v=vs.110\).aspx](https://msdn.microsoft.com/ru-ru/library/hh156542(v=vs.110).aspx) (дата обращения 11.02.2016).
21. Sriram Krishnan. "Programming Windows Azure: Programming the Microsoft Cloud". 2010. 368 с.



22. Рихтер Дж. “CLR via C#. Программирование на платформе Microsoft .NET Framework 4.0 на языке C#”. 3-е изд. - СПб.: Питер, 2012.- 928 с.
23. Paul Viola, Michael Jones. Rapid object detection using a boosted cascade of simple features // Accepted conference on computer vision and pattern recognition, 2001.
24. VladimirLeskin/Facial-emotion-recognition // GitHub. URL:  
<https://github.com/VladimirLeskin/Facial-emotion-recognition> (дата обращения 13.04.2016).
25. About Git // Git. URL: <https://git-scm.com/about> (дата обращения 13.04.2016).
26. EmoRec – приложение ASP.NET // azurewebsites. URL:  
<http://cameracaptureleskinv.azurewebsites.net/> (дата обращения 13.04.2016).